

The University of Chicago

THE CONSTRUCTION OF SOCIAL ATTITUDE SCALES

A PART OF A DISSERTATION SUBMITTED
TO THE FACULTY OF THE DIVISION OF
THE SOCIAL SCIENCES IN CANDIDACY FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATION

1933

By

ARLYN CUSTER ROSANDER

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

1933

The University of Chicago

THE CONSTRUCTION OF SOCIAL ATTITUDE SCALES

A PART OF A DISSERTATION SUBMITTED
TO THE FACULTY OF THE DIVISION OF
THE SOCIAL SCIENCES IN CANDIDACY FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATION

1933

By

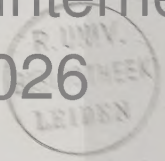
ARLYN CUSTER ROSANDER

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

1933



Digitized by the Internet Archive
in 2026



https://archive.org/details/IA41533907_0114

THE CONSTRUCTION OF SOCIAL ATTITUDE SCALES

The purpose of this investigation was to improve the methods of constructing social attitude scales by evaluating some of the major assumptions, by simplifying present techniques, by formulating new criteria, and by designing new procedures in this field of scale construction.

By social attitude scale we mean a special paper and pencil test in which the individual indicates agreement or disagreement toward each element of an ordered series of stimuli which may be either opinions toward social issues or responses to life situations. This is the type of scale which Allport and Hartman employed several years ago, and which Thurstone has refined to a considerable extent. We present herewith the results of our studies with regard to the optimum number of sorters, the qualifications of the sorters, the different methods of scaling and indexing, the reliability of an attitude scale, the validity of an attitude scale, and the magnitude of various errors arising in the process of construction.

The optimum number of sorters.- Fifty carefully selected statements of opinion toward the issue of German war guilt were sorted by 50 college students, 28 undergraduates and 22 graduate students, into 11 piles according to the usual Thurstone technique. These statements were divided into two parallel groups, and ranked by the same group of sorters. The Spearman-Brown formula was applied to the data from the sorted statements and to those from the ranked statements for groups of sorters of 2, 3, 4, 5, 10, 15, 20 and 25 in number. The computed and the predict-

ed correlation coefficients were not significantly different for either set of data. The reliability coefficient of 25 sorters of the type employed can be safely placed at .995; even 15 sorters give a coefficient of .990. In addition, 30 church statements from Thurstone and Chave's master list, statements which had been scaled by 300 sorters, were re-scaled by using only 25 judges. The correlation coefficient between the two sets of scale values was .988 with no significant differences between either the means or the standard deviations. More than 25 sorters of this type are hardly justified in ordinary social attitude construction.

The qualifications of the sorters.- In this part of the study we compared graduate and undergraduate sorters, different groups of subject matter specialists, experienced sorters with inexperienced sorters, biased with unbiased sorters. Using the extent of correlation of the individual judgments with those of the entire group as the criterion of satisfactory sorting, we found no significant difference between graduate and undergraduate sorters, the group of sorters most highly acquainted with the issue of war guilt came out poorly, but the experienced sorters were better than the average of the group. Both the coefficient of internal agreement and factor analysis showed no significant differences between graduate and undergraduate sorters considered as groups. Three groups of biased sorters were used to scale 68 prohibition statements by the usual sorting method. By means of the Thurstone-Smith prohibition scale we selected one group of wets, one group of neutrals, and one group of drys. The scale values obtained from the sortings of each group were correlated with those of the other groups; the values ranged from .975 to .985. In addition, the sortings of each group were analyzed; a number of sorters were carefully observed while they were making

their judgments. Some extremely biased sorters agreed or disagreed with the statements instead of judging them according to the degree of favorableness or unfavorableness which they expressed. Extremists did not scale their own views as extreme as their opponents scaled them. All in all, it was found that the sorters with a score in the middle of the scale were preferable to those whose scores were at the extreme ends of the scale. Sorters with very extreme views toward the issue which they are scaling should be avoided.

Methods of scaling and indexing.- Two methods of scaling and two methods of indexing were compared; the methods of equal appearing intervals (e.a.i.) and equally often noticed differences, and the methods of mean rank order and median rank order. The elements used were the 50 statements of the two forms of the German war guilt scale. The mean rank orders and the median rank orders for these statements gave a plot that was perfectly linear for all practical purposes. The following correlation coefficients were obtained between statement values obtained by these four methods: that between the e.a.i. method and the equally often noticed difference or proportion method was .984, that between proportions method and mean rank order was .989, that between the e.a.i. method and mean rank order was .995, and that between e.a.i. and median rank order was .995. These correlation coefficients indicate a high degree of statistical equivalence for the sets of scale values obtained by these methods.

We showed that the end effect inherent in the method of sorting might be greater at one end of the scale than at the other, and pointed out how this distortion might be rectified by means of corrected e.a.i. values derived from an empirical e.a.i.-proportions curve.

We analyzed the two methods of scaling mentioned above and showed that even though the dispersions of the judgments may be equal, the results may not be identical. There is no reason why they should be identical inasmuch as the e.a.i. method is based upon the median judgment along a series of piles, whereas the proportions method is based upon the sum of the individual judgments of stimulus A greater than stimulus B.

The reliability of an attitude scale.- A social attitude scale based upon reactions to behavior situations was found to be highly reliable but not so reliable as one based upon statements of opinion. The value of this coefficient for 22 Negro statements given to approximately one hundred northern college students was .899; a behavior situation scale of 44 elements on the same issue given to the same group gave a reliability coefficient of .856. On the Negro opinion scale the southern college students tested gave a reliability coefficient somewhat lower than that obtained from the northern students. The same tendency was true of the behavior situation scale; the reliability coefficient for northern students was .813 while that for the southern students was .736, after making a correction for differences in range of scores.

The reliability of a scale is influenced by the method of scoring. One might use either the median or the mean of the scale values of all the statements or elements which an individual checks. The former is much easier to compute; it is not so accurate. The correlation coefficient between the scores of 100 Negro scales each scored by both methods was .98, which indicates a high degree of equivalence.

It is possible to obtain a high reliability coefficient between parallel forms of an attitude indicator by careful selection of the statements and by repeated ranking of them by compe-

tent judges. The value of each statement may be expressed by its rank in the series providing the statements are fairly well graded; this gives us a kind of quasi-scale. We constructed two of these quasi-scales, one on the League of Nations and one on German war guilt. The reliability coefficient of 31 statements of the former scale, based upon 200 cases was .915. The reliability coefficient of 25 German war guilt statements, based upon 200 cases, was .924. This indicates that a high reliability coefficient may be obtained by this simple method of construction. It is not recommended, however, for those who are inexperienced in attitude scale construction.

The instability of the individual's attitude was measured by the per cent of identical responses on two successive applications of the same scale, precautions being taken to prevent memory effects. For a group of 60 students this coefficient of instability was found to be 74% with respect to the Negro opinion scale, 74% for one form of the German war guilt scale, and 73% for the other form. The instability of response has quite an appreciable effect upon the repeat reliability coefficient; it deserves to be much more thoroughly investigated.

The validity of an attitude scale.- Two forms of the Negro opinion scale and two of the Negro behavior situation scale were administered to the same group of about 100 college students, the two sets were correlated and the resulting values corrected for attenuation. The corrected correlation coefficient for one form was .812, for the other it was .891; the average corrected correlation was .851. This indicates a high degree of commonality between these two types of attitude scales.

Further studies were made of the soundness of the techniques used in the process of constructing social attitude scales.

Thirty-eight statements from the master list of communism statements collected by Thurstone, which had previously been scaled by the e.a.i. or sorting method were re-scaled by 40 judges. They arrived at the scale value of each statement by comparing it with those statements in the communism scale already constructed. The scale values found by this method of equivalent statements correlated with those obtained by sorting to the extent of .98.

Using the coefficient of discrimination which is the correlation between scale values of statements and scale scores of individuals for a given scale, we increased the discriminating power of the Negro opinion scale by reducing the number of statements from 22 to 11. We did this by making use of the distribution of endorsement proportions of every statement; we also made use of the instability ratio which is the ratio of the number of times a statement was checked once to the number of times it was checked twice for a scale administered on two successive occasions. In this way we increased the coefficient of discrimination of the Negro scale from .644 to .713.

We concluded that the coefficient of agreement among sorters, the correlation between scale values of statements and the position ranked or sorted should be at least .90. In addition, we estimated that the coefficient of discrimination for a given set of statements administered to an adequate number of individuals should be at least .70.

The magnitude of the various errors.- Two methods never lead to exactly the same results; a given task is not performed exactly the same this week that it will be next week. Particularly is this true of social attitude scale construction where we deal with a functional, and not a structural phenomenon. The difference between two sets of values obtained for the same specific

purpose expressed in terms of 11 divisions or 25 ranks constitutes the per cent error arising from using one method rather than the other. We give below a number of these.

1. Mean versus median method of scoring.....	3.3%
2. Group of 25 sorters versus group of 300 sorters.....	3.6
3. Instability of individual response on two successive occasions	
Negro scale.....	4.6
War guilt scale, Form A.....	4.2
War guilt scale, Form B.....	3.6
4. Inconsistency of sorters on two successive occasions	
sorting.....	7.1
ranking.....	7.9

These data show that the errors due to the instability of the individual response and the inconsistency of the sorters are of a much greater magnitude than those due to scoring methods or the number of sorters, issues in which we have been most interested up to the present time.

Needed investigations.- The construction of several more scales based upon the behavior situation would be desirable. New scales employing pictorial elements such as cartoons would be of great value. Someone might attempt to scale elements of experience such as occurred in Stouffer's prohibition histories and which formed the cues to the judgments of the raters.

Improvement of social attitude scales already constructed will be necessary as we discover new methods of refinement. We need to apply more criteria, or better criteria, in the selection of the final statements of the scale.

Analysis of attitude scales is needed in order to determine what common factor or factors, if any, underlie the several attitude variables for which scales have been constructed.

An application of great significance is that of determining the effectiveness of teaching in changing the social attitudes of students. Numerous experiments should be set up to investigate

this problem.

With improved scales, we shall be in a position to determine the rise and shift of social attitudes in elementary and secondary school students.

We need to develop scales which are much more specific and diagnostic; perhaps the behavior situation scale is a step in this direction.

